Human Interest or Hard Numbers?

Experiments on Citizens' Selection, Exposure, and Recall of Performance Information

Olsen, Asmus Leth

Published in:
Public Administration Review

DOI:
10.1111/puar.12638

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

PAR PUBLIC ADMINISTRATION REVIEW

Asmus Leth Olsen
University of Copenhagen, Denmark

Human Interest or Hard Numbers? Experiments on Citizens' Selection, Exposure, and Recall of Performance Information

Asmus Leth Olsen is associate professor in the Department of Political Science at the University of Copenhagen, Denmark. His research lies in the areas of behavioral public administration, political-administrative psychology, performance management, and experimental public administration. His current research focuses on the effects of performance information on citizens and the role of dishonesty for selection into public sector employment. His work has appeared in *Political Behavior, Public Administration Review,* and *Judgment and Decision Making.*
E-mail: ajlo@ifs.ku.dk

Abstract: *The abundance of quantitative performance information has motivated multiple studies about how citizens make sense of "hard" performance data. However, research in psychology emphasizes that episodic information (e.g., case stories) often leaves a greater mark on citizens. This contradiction is tested using multiple experiments embedded in a large, nationally representative sample of Danish citizens. The results stress three differences between statistical and episodic data. Citizens have strong preferences for statistical data when asked to evaluate an organization. However, episodic information has in some instances a stronger impact on citizens' evaluations of an organization and often is more emotionally engaging than statistics. Finally, when asked to immediately recall recent performance information about public services, citizens report more elaborate information about personalized stories and experiences than about statistics. Overall, the results raise questions about the ability of hard performance data to dominate and crowd out episodic performance information.*

Practitioner Points

- Providing citizens with quantitative performance data will not automatically crowd out their reliance on personal experience or media case stories.
- Episodic experiences with public services provide a more vivid account of performance, which is more emotionally engaging and easier to recall for citizens.
- Policy makers and managers must pay closer attention to how their performance data can provide citizens with a more vivid and emotional account of public services to complement the often pallid statistical performance data.

Performance measurement is fundamentally about assigning numbers to the inputs, outputs, and outcomes of public organizations. Accordingly, when we refer to performance information, we imply *quantitative* performance information. Increasingly, performance data are intended for the public at large. This trend has sparked a natural research interest in how citizens make sense of such numbers. Key experimental findings show that performance data clearly affect citizens' attitudes (Bækgaard 2015; James 2011). Moreover, citizens draw on reference points for comparisons (Charbonneau and Van Ryzin 2015; Hansen, Olsen, and Bech 2015; James and Moseley 2014; Olsen, forthcoming), and they engage in motivated reasoning and draw on implicit attitudes in their interpretations of performance data (Bækgaard and Serritzlew 2016; Marvel 2016). In addition, simple framing effects can alter the inferences that citizens draw from the data (James and Van Ryzin 2015; Olsen 2013, 2015a).

On the other hand, while the abundance of public performance data motivates these studies, we have overlooked the fact that "hard" performance data is only a subset of the information about public services that is available to citizens. In fact, empirical studies show that the dominant source of information for citizens is often the exact opposite of statistical numbers: "episodic" information in the form of informal "human interest" stories from personal experience, word-of-mouth accounts from friends and family, and case story coverage of public services in the media (Grosso and Van Ryzin 2011; Kettl 2016; Slattery and Hakanen 1994). That is, even in the age of quantitative performance information, citizens are still exposed to their neighbors' experiences from the hospital or news stories about misconduct by a teacher at the local school.

At the managerial level, it has also been noted that episodic information attracts considerable attention (Bolman and Deal 1991; Kroll 2013; Mintzberg 1971; Moynihan 2008). This literature notes that managers often draw on "soft" information in the form of case stories from the news, ad hoc meetings, and casual observations. For instance, Mintzberg

The copyright line for this article was changed on 26 December 2017 after original online publication.

Public Administration Review, Vol. 77, Iss. 3, pp. 408–420.

DOI: 10.1111/puar.12638.

argues that managers "favor verbal channels, face-to-face contact and telephone and, to a lesser extent, direct observation as means of supplementing (and often replacing) formal sources of information" (1973, 3). In psychology, there has been a long-standing interest in contrasting the effect of such "episodic" information with more abstract statistical information (Nisbett and Ross 1980). Here, the evidence suggests that episodic information is more potent than statistical information on multiple dimensions. For one, episodic information is generally more vivid, and vividness is an important determinant of evaluative impact (Nisbett and Ross 1980; Pettus and Diener 1977). Second, episodic information is more emotionally engaging than abstract numbers (Slovic 2007). Moreover, episodic information is more easily stored in and retrieved from memory (Aarøe and Petersen 2013; Nisbett and Ross 1980). Taken together, existing work outside of performance management on episodic information calls into question the relative effectiveness of numerical performance information.

This article directly confronts this challenge and undertakes the first experimental test comparing how citizens rely on episodic and statistical information about public sector performance. The test of episodic and statistical data focuses on three distinct dimensions: selection, exposure, and recall. First, it investigates citizens' *stated preferences* for either statistical or episodic information when asked to evaluate a service. Second, it tests how episodic and statistical information changes *evaluations* of public services and induces different emotional responses. Third, it estimates the probability of *recalling* episodic and statistical performance information and observe differences in how elaborate the recalled information is.

In order to do so, the analysis relies on a set of experiments embedded in a single survey. The survey is fielded in a large, nationally representative sample of the Danish population ($n = 1{,}013$). A nationally representative sample provides strong external validity in terms of generalizing the experimental findings to the broader public. At the same time, the experiments are simple and general enough to allow for replication and extension in different contexts and countries.

The findings show that numerical performance information, across many dimensions, is less influential than episodic descriptions of individual experiences with public services. On one hand, citizens have stronger stated preferences for statistical performance information, but apart from explicit information selections, we find that episodic performance information tends to be more potent. Citizens' evaluations of public services are more strongly affected by episodic information, and if people are faced with negative information, it also elicits a stronger emotional response in terms of "compassion." In terms of recall, citizens are more likely to recall episodic information, and their memories are more elaborate regarding episodic information instead of statistical information. In short, episodic performance information affects human minds and sticks.

Citizens' evaluations of public services are more strongly affected by episodic information, and if people are faced with negative information, it also elicits a stronger emotional response in terms of "compassion."

The findings of this study have implications for future research and practitioners' views of performance data. First, the findings indicate that although numerical performance information diffuses across countries, sectors, and levels of government, it will not automatically crowd out existing sources of episodic information. As practitioners and researchers, we have a task ahead of us in trying to understand how to balance human interest and hard numbers when communicating performance information to the public. Second, while the focus is on the effect of episodic and statistical information at the *citizen level,* the findings have implications for how performance information may affect managers and employees as well. These implications are discussed in the concluding section.

How Statistical Performance Information Affects Citizens

Most definitions of performance measurement or performance information stress the importance of assigning numbers to the inputs, outputs, and outcomes of public organizations (Davies 1999; Hood 2007; James 2011; Moynihan 2008; Perrin 1998; Van Dooren, Bouckaert, and Halligan 2010). As noted by Radin, "There is perhaps no element within the performance measurement process that is more important than the reliance on numbers and quantitative presentation of accomplishments" (2006, 27). Accordingly, the numerical aspect has often been stressed as a potent driver that affects the perceptions, evaluations, and behavior of managers, employees, and citizens. At the organizational and managerial levels, this sentiment is captured in the notion of "what gets measured gets done" (Osborne and Gaebler 1992, 146). It implicitly conveys the idea that performance information is able to crowd out other sources of unmeasured information about public services. Holzer and Yang note that "measurement helps to move the basis of decision making from personal experience to proof of measurable accomplishment or lack thereof" (2004, 16). However, numbers have also been seen as forces for unintended, albeit still powerful, consequences. For instance, Smith raises concerns about "tunnel vision," which takes the center stage of attention "at the expense of unquantified aspects of performance" (1995, 284). In other words, numbers have the potential to crowd out the episodic pieces of information that remain uncounted.

What could explain the fact that quantitative performance information is successful in affecting citizens? One view is that the assignment of numbers to complex social outcomes has the advantage of absorbing potential uncertainties and diverging signals of the underlying data (March and Simon 1958; Moynihan 2008; Stone 2012). Espeland and Sauder mention how quantitative "information appears more robust and definitive than it would if presented in more complicated forms" (2007, 17), and Jackson adds that "ambiguous concepts, however, make deceptive statistics" (2011, 23). The experimental work on how performance information affects citizens has also stressed the heuristic qualities of performance data: it helps simplify the otherwise complex task of evaluating whether a public organization is performing well or poorly (James 2011; James and Moseley 2014). Quantification also hints at an implicit cue about the importance of the information, as only important matters are worth counting (Stone 2012, 192). An

extension of this argument stresses the scientific cues of performance numbers: "Numbers generated by sophisticated statistical techniques can give a false sense of objectivity and reliability" (Jackson 2011, 24). In contrast to anecdotes or personal experience, numbers seem more objective and have a more decisive nature (Herbst 1993).

From social psychology, we know that precision—even when unwarranted—can be very persuasive because it carries implicit cues about the knowledge of the communicator (Zhang and Schwarz 2013). In other words, if we assign a number to the performance of an organization, we communicate to the world that we have highly accurate information about that organization's performance (Olsen 2015b).[1] Stone also notes how "numbers, by seeming to be so precise, help bolster authority of those who count" (2012, 196). Taken together, we encounter a view in which numbers seem objective, unambiguous, and precise to the receiver of the information. We predict that these qualities are important when citizens engage in *explicit* information selection. That is, cues about simplicity, objectivity, importance, and accuracy will make statistics more appealing in direct comparison with episodic information, which usually scores lower on these dimensions. This realization leads us to our first hypothesis:

> **Hypothesis 1**: When making deliberate decisions about what information to engage with, citizens are more likely to select statistical information over episodic information.

However, as we will discuss in the following section, there are good reasons to believe that statistical performance information—for citizens—will be less persuasive in actual evaluations, less emotionally engaging, and more difficult to recall.

How Episodic Performance Information Affects Citizens

There has been little to no systematic attempt to compare the effects of episodic and statistical performance information. Even so, we can trace a concern about the ability of numbers to impact citizens in the very early works on performance reporting. Clarence Ridley is well known for his extensive work on municipal and city reporting from the 1920s and onward. He had an eye for the attractiveness of vivid episodic information and called for the importance of "human interest treatment of other-wise dull factual data" (Ridley 1937, 113). Others were also fundamentally concerned with the ability of statistics to engage citizens and noted how most city reports were "dull, statistics-impregnated affairs" (Roher 1941, 196). However, while the fundamental division between the two forms of information exists and dates back to the early days of performance management, we have to turn to communication studies and political psychology to find studies that test and compare their impact on citizens.

In these neighboring fields, the contrast to numbers is reflected in the focus on "episodic" information versus "thematic" (e.g., statistical or abstract) information.[2] Communication research has a natural interest in episodic information because cases and exemplars are central to modern mass-media journalism. The main difference between episodic and statistical information lies in their degree of vividness (Pettus and Diener 1977). Episodic information is highly vivid, for it draws on "personalized case histories," while statistical information deals with "abstract concepts and general trends" (Iyengar 1990). Vivid information is high on "concreteness," which includes the "degree of detail and specificity about actor, actions, and situational context" (Nisbett and Ross 1980, 47). Moreover, vivid information means imaginable information that has a "tendency to prompt sensory information" (Nisbett and Ross 1980, 47). Across multiple studies, such "image-provoking" episodic information has been found to be more persuasive than statistical information (Herr, Kardes, and Kim 1991; Pettus and Diener 1977; Zillmann 2006). For instance, Daschmann experimentally compares the effect of vox pop information and a public opinion poll and finds that vox pop information is "considerably stronger, to the point of overriding the effects of poll results" (2000, 160). Importantly, an overreliance on informational vividness is to be considered a bias because vividness can be orthogonal to importance, relevance, or some other dimension that individuals should care about (Nisbett and Ross 1980, 62). Drawing on the existing body of evidence regarding the effects of vivid information provides a clear expectation as to how citizens are affected when exposed to either episodic and statistical information:

> **Hypothesis 2**: Episodic performance information will have a stronger effect on evaluations of public services than statistical information describing the same aspect of performance.

Another dimension of episodic information is the ability to evoke affective and emotional responses beyond what statistical information can achieve (Gross 2008). Indeed, vivid information is often seen as more "emotionally interesting" (Nisbett and Ross 1980, 45). As an example, in political communication, Aarøe (2011) finds that an episodic framing of policy changes to immigration elicited stronger emotional responses in terms of compassion, pity, anger, and disgust. We also find a related argument in studies on how statistical information induces "psychic numbing" and "compassion fade" (Jenni and Loewenstein 1997). As Slovic notes, "Numbers fail to spark emotion or feeling and thus fail to motivate action" (2007, 80). For instance, we are less willing to save "statistical victims" because we find it difficult to be emotionally engaged with their suffering and pain. In contrast, we often see an "identifiable victim effect," which denotes an enhanced willingness to spend more money on saving the lives of identifiable victims rather than statistical lives (Kogut and Ritov 2005). We expect to find a similar difference in the emotional arousal of statistical and episodic information:

> **Hypothesis 3**: Episodic information about performance is more emotionally engaging than statistical information.

The final dimension is the recall of episodic and statistical information from memory. For Nisbett and Ross (1980), an important aspect of episodic information's potency is that it is more likely to be recalled. Ease of recall is consequently part of the explanation as to why vivid information can have a

Ease of recall is consequently part of the explanation as to why vivid information can have a disproportional effect on judgment and evaluations: memorable information will be given more weight in judgment.

disproportional effect on judgment and evaluations: memorable information will be given more weight in judgment.[3] At the same time, the fact that vivid information can be more image provoking and elicit stronger emotional responses makes it more memorable (Taylor and Thomson 1982, 162).[4] A special case involves instances from our own lives, which will be even more easily recalled and with much greater detail—often referred to as "episodic memory" (Tulving 2002). Statistical information, on the other hand, is marked by the complete opposite characteristics and will be less likely to be stored in memory and thus less likely to be retrieved for later evaluation. In fact, it has been noted in performance management research that numerical performance information may be much more difficult to remember (Grosso and Van Ryzin 2011; Lee 2006). For the same reason, early studies of performance data argued for the relevance of including photographs and other colorful material in order to make the data more memorable (Roher 1941; Upson 1915). Along the same lines, we expect a difference in citizens' ability to recall episodic and statistical information:

> **Hypothesis 4**: Episodic information is more easily recalled and also recalled with greater nuance than statistical information about public sector performance.

Data and Design: Three Experimental Studies in a Nationally Representative Sample

The four hypotheses are tested with three experiments using a sample of citizens representative of the Danish population. In study 1, the focus is on citizens' stated preferences for either statistical or episodic performance information when evaluating a public service. This allows us to test whether statistical information is in fact preferred over episodic information in explicit choice situations (hypothesis 1). Study 2 is devoted to the differences in how statistical and episodic performance information affect citizens' evaluation of a service (hypothesis 2) and their emotional responses to both types of information (hypothesis 3). Finally, study 3 addresses differences in the probability and degree of recall concerning episodic and statistical information (hypothesis 4).

All three experiments were embedded in a large, nationally representative sample recruited through YouGov's Danish online panel ($n = 1{,}013$). Subjects were randomly assigned to different experiments and different treatment conditions within the survey. The data were collected in late May 2015 over the course of five days.[5] The sample was restricted to citizens between the ages of 18 and 74 and prestratified on gender, age, geographic region, and political party choice in order to achieve a near-representative sample of the Danish adult population (50.0 percent female, mean age of 46.6 years, and 32.0 percent residing in the capital region).[6] By applying a large representative sample of the general population, we can ensure that any results obtained are externally valid to the Danish public at large. Furthermore, by applying a nationally representative sample, the results are more easily compared with experimental evidence obtained in other contexts and settings.

Study 1: Citizens' Preferences for Statistical and Episodic Performance Information

The purpose of the first experiment is to determine whether citizens prefer episodic or statistical information when asked to rate a public service (hypothesis 1). The focus is therefore on active selection of performance information rather than the effects of exposure.

Table 1 Experimental Conditions

Question for All Subjects:		
Imagine that you have to evaluate how good or poor a hospital is. You can choose one of the two following pieces of information. What would you choose?		
Treatment Conditions (randomly assigned):		
Statistical vs. Case of "Former Patient" ($n = 102$)	Statistical vs. Case of "Birgitte" ($n = 98$)	Statistical vs. Case of "Erik" ($n = 107$)
Results from a satisfaction report with 100 former patients	*Results from a satisfaction report with 100 former patients*	*Results from a satisfaction report with 100 former patients*
The experience of a former patient	*The experience that 51-year-old Birgitte had as a patient*	*The experience that 31-year-old Erik had as a patient*

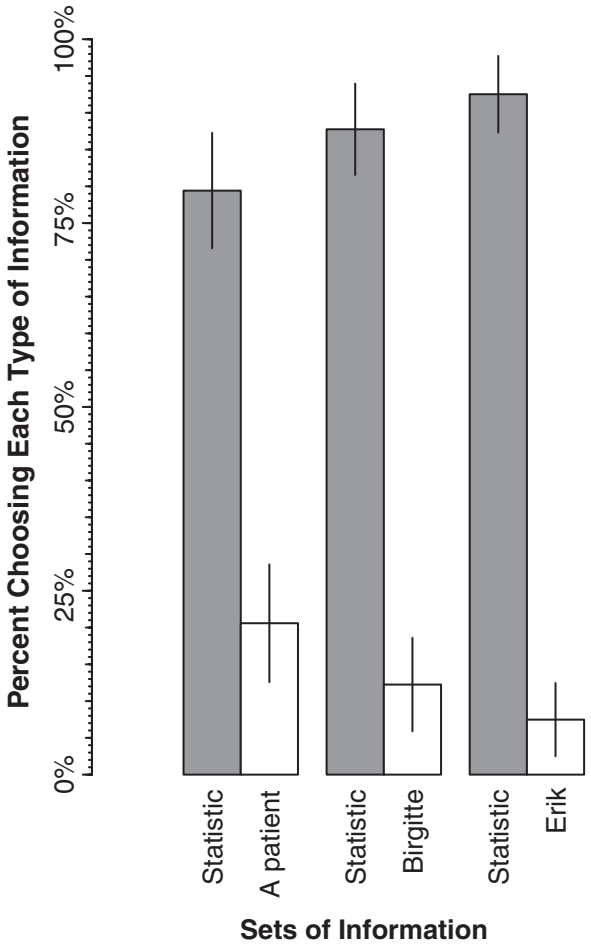
Experimental Design and Procedure

The experiment is a simple between-subjects design with three conditions ($n = 307$). All subjects are asked to imagine that they have to evaluate the performance of a hospital. In order to do so, they are permitted to choose one of two distinct pieces of information. In all three conditions, one of the two pieces of information is a report on the satisfaction of 100 former patients. It represents a typical form of numerical performance data that is usually available to citizens. Across the three conditions, the alternative information is one of three case stories representing episodic data. One story is about an unnamed former patient, while the two other stories involve an identifiable older female ("Birgitte") and a younger male ("Erik"). The conditions are described in table 1.

In the experiment, the different forms of episodic conditions are employed in order to verify whether potential differences are robust across very diverse types of episodic information. An example of one of the treatment conditions, as it was shown to the subjects, is reported in appendix A. In addition, the order of presentation of the two options is randomized for each subject in order to avoid sequence effects and keep from implicitly indicating to the subjects that one piece of information is more important than the other. Subjects had to choose one of the two options in order to proceed in the survey. Median response time for the question was 15 seconds. The experiment is not necessarily intended to be the optimal way of capturing the type of information that citizens *actually* draw on. Instead, the purpose is to capture what form of performance information citizens prefer when asked to make an explicit information selection.

Empirical Results

The results of this experiment are reported in figure 1. Across all three conditions, we find that citizens strongly prefer statistical information from the satisfaction survey over the episodic description of individual patients. For the condition with an unidentifiable former patient, 79.4 percent prefer statistical information ($p < .001$). For the condition with a 51-year-old woman ("Birgitte"), around 87.8 percent prefer statistical information ($p < .001$), and in the 31-year-old male condition ("Erik"), the same estimate is 92.5 percent ($p < .001$). If we combine all three



Note: Percentage of respondents choosing each of the two types of information in the three different conditions. Lines constitute 95 percent confidence intervals (n = 307).

Figure 1 Citizens' Choice of Either Statistical or Episodic Performance Information

conditions, then 86.6 percent (confidence interval = 82.2 percent to 90 percent) of the subjects chose the statistical information. Overall, this result provides strong evidence of citizens' stated preference for statistical information over single case stories when directly asked to choose between both types information about a public service (hypothesis 1).

Study 2: Evaluative and Emotional Responses to Statistical and Episodic Information

Study 1 is a simple test of citizens' stated preferences for either episodic or statistical performance information regarding public services. In study 2, we flip the question and ask how citizens' evaluations of public services are affected by either episodic or statistical performance information (hypothesis 2). We also look at the differences in their emotional responses (hypothesis 3) when exposed to either statistical or episodic information.

Experimental Design and Procedure

The experiment is a between-subjects design with five conditions (n = 706). In line with the existing research, our primary test of vividness is done by comparing the effects of the three episodic treatments with the effects of the two statistical treatments (Kogut and Ritov 2005; Nisbett and Ross 1980). All subjects are provided a piece of performance information about an unnamed hospital, as shown in table 2. The experimental conditions reflect different segments of information that subjects can choose in study 1. We hereby receive a very direct comparison between differences in information *selection* and the *effects* of randomly assigned information exposure. In order to avoid spillover effects, subjects from study 1 were excluded from participating in study 2. This is because prior knowledge of similar information would most likely influence subsequent exposure effects.

All five experimental conditions focus on instances of a simple surgery performed by a hospital. Hereafter follows a description in which a single individual or a sample of patients experience serious pain because of an erroneous hospital treatment. Three of the conditions are simple case examples with one unidentifiable "patient" and two identifiable cases of "Birgitte" and "Erik" (as used in study 1). The two identifiable cases are intended to increase the vividness of the information by providing concrete details with names (and thus indirectly gender) and age. This is a conservative manipulation of vividness that draws on similar tests conducted in psychological research on identifiable victims (Kogut and Ritov 2005, 161; Slovic 2007). The treatment conditions are much more subtle than those found in recent studies of episodic framing in political psychology. Such cases usually implement colorful and emotionally charged language (Aarøe 2011; Gross 2008).

Table 2 Experimental Conditions

Information Provided to All Respondents:				
Imagine that you know the following about a hospital . . .				
Treatment Conditions (randomly assigned):				
Episodic I: "A Patient" (n = 142)	Episodic II: "Birgitte" (n = 144)	Episodic III: "Erik" (n = 137)	Statistical I: "1%" (n = 139)	Statistical II: "10%" (n = 144)
A patient was admitted to the hospital last year in order to have a simple surgery.	*51-year-old Birgitte was admitted to the hospital last year in order to have a simple surgery.*	*31-year-old Erik was admitted to the hospital last year in order to have a simple surgery done.*	*100 patients were admitted to the hospital last year in order to have a simple surgery.*	*100 patients were admitted to the hospital last year in order to have a simple surgery.*
Following the surgery, the patient was in great pain due to erroneous treatment.	*Following the surgery, Birgitte was in great pain due to erroneous treatment.*	*Following the surgery, Erik was in great pain due to erroneous treatment.*	*Following the surgery, 1% of the patients were in great pain due to erroneous treatment.*	*Following the surgery, 10% of the patients were in great pain due to erroneous treatment.*
How good do you think the hospital is?				
On the Following Page:				
To what extent does this information make you feel: compassion, disgust, anger				

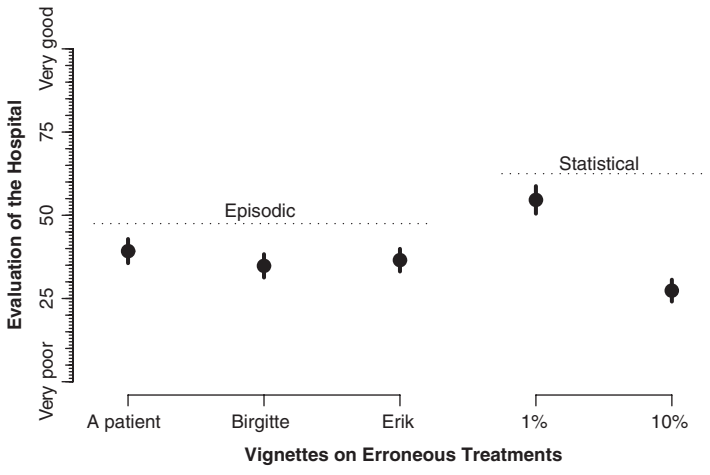
The two final conditions present statistical information for 100 patients from the hospital with varying percentages of them experiencing an erroneous hospital treatment (1 percent or 10 percent). The differing degrees of failure in the two statistical treatments provide an indication of how robust the results are if episodic information of an erroneous treatment is compared with both "mild" (1 percent) and "severe" (10 percent) statistical evidence of erroneous treatment. Formally, the "1 percent" treatment also only describes erroneous treatment of a single individual (1 patient out of 100). Therefore, it is our primary benchmark for a direct comparison of the impact of our three episodic treatments that also describe the experience of a single individual (hypothesis 2). Across all conditions, the descriptions of the surgery and the erroneous treatment are held constant in order to only manipulate the form of information (i.e., episodic/statistical) and not the actual content. An example of one of the treatment conditions, as they were shown to the subjects, is reported in appendix B.

Beneath the various erroneous treatment descriptions, subjects are asked to provide an evaluation of the hospital in question. The evaluation is our dependent variable and is provided on a slider scale ranging from 0 ("very poor") to 100 ("very good"), with no possibility of not providing a response. Across all conditions, the average evaluation is 38.4 (SD = 23.5). Median response time was about 17 seconds. The low average evaluation is expected given the negative emphasis on "great pain" and "erroneous treatment" across all conditions. It underlines that in terms of negative information, the experiment was successful in manipulating a negative view of the hospital.

On the following page, all subjects are asked the extent to which the information induced three distinct emotional responses, namely "compassion," "disgust," and "anger." These emotional responses have been applied in similar studies of episodic and statistical information (Aarøe 2011; Gross 2008). The three emotions are presented in random order for each subject in order to cancel out any sequence or anchoring effects between emotions. Subjects rate the extent to which they feel the three emotions on a 11-point scale ranging from "not at all" (0) to "very much" (10). Mean evaluations across all conditions indicate that all three emotional responses have been activated: compassion (M = 7.1, SD = 2.6), disgust (M = 6.0, SD = 3.0), and anger (M = 5.4, SD = 3.1). The median response time was about 16.5 seconds. Again, the relative high scores for the three negative emotions confirm that, on average, the negative wording of the treatment conditions has a negative emotional impact on subjects.

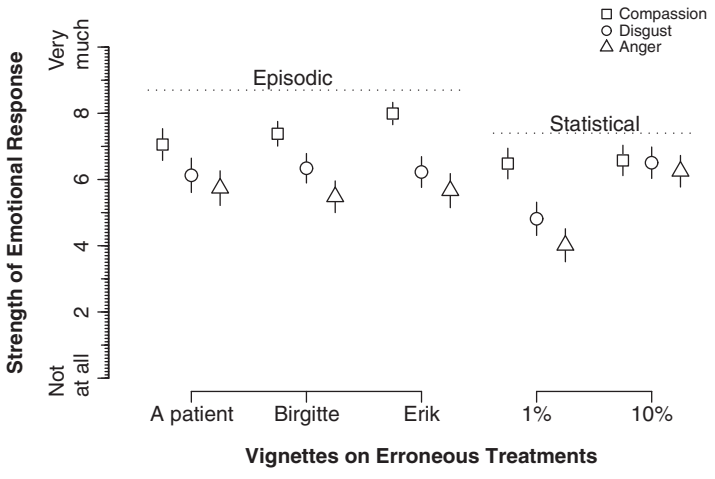
Empirical Results

The findings from study 2 on overall hospital evaluation are reported in figure 2. Here are shown the mean evaluations with 95 apercent confidence intervals for all treatment conditions. The figure shows great variation in average evaluations between the episodic and statistical conditions. Each of the three episodic conditions causes significantly and substantially lower evaluations than the 1 percent statistical condition, which receives the most favorable rating of 54.6 points (SD = 24.8). If we combine the three single cases and test directly against the 1 percent erroneous treatment condition, we obtain an estimate of a 17.8 point lower evaluation ($p < .001$). The condition with a severe failure rate of 10 percent receives the lowest evaluation of 27.4 (SD = 20.1), which is significantly lower than all of the individual single cases and in combination with an estimated difference of 9.5 points ($p < .01$). However, if we create a combined test, the three single cases receive a lower evaluation than the two statistical conditions (3.9 point difference, $p < .05$). Finally, we note that the two identifiable cases receive a slightly lower evaluation than the unidentifiable one—although the difference is not significant or substantial of magnitude (3.6 points, $p = .11$). Overall, we note that episodic information has a substantially stronger negative impact on citizens' evaluations of public services than statistical information of 1 percent erroneous treatment rate out of 100 patients (hypothesis 2). Objectively, the statistical information also described the experience of a *single*



Note: Dots represent mean evaluations of hospital performance for each experimental condition. Lines constitute 95 percent confidence intervals (n = 706).

Figure 2 Citizens' Evaluations of a Hospital across Five Experimental Conditions Describing Erroneous Treatment at a Hospital



Note: Dots represent mean emotional responses for each experimental condition. Lines constitute 95 percent confidence intervals (n = 706).

Figure 3 Citizens' Emotional Responses across Five Experimental Conditions Describing Erroneous Treatment at a Hospital

individual (1 percent out of 100 patients), but it did not have remotely the same negative impact on evaluations as the episodic case descriptions. However, this does not imply that episodic information has a greater impact than all levels of statistical evidence on the same problem. As an important caveat, we note that if the statistical information indicates severe performance problems (i.e., 10 percent failure rate), then it does have a slightly stronger negative impact.

Next we turn to the question of how the conditions affected citizens' emotional responses. These results are shown in figure 3. Overall, the results highlight great variation across emotional responses between the five conditions. The most striking difference is the substantively and significantly higher level of compassion across the three single cases when compared with both of the statistical conditions. The difference amounts to about 1 point on the 11-point scale ($p < .001$). For disgust, the combined difference is smaller at 0.56 point ($p < .05$) and for anger at 0.47 point ($p < .05$). However, for disgust and anger, the emotional difference is mainly driven by the 1 percent erroneous treatment condition. Generally, we find less of an effect for disgust and anger when we directly compare each of the statistical conditions with the combination of the three episodic ones. For compassion, the episodic cases have a significantly and substantively greater impact than both the 1 percent condition (0.99 point, $p < .001$) and the 10 percent condition (0.89 point, $p < .001$). However, for disgust, there is only a difference if the episodic treatments are compared with the 1 percent condition (1.42 points, $p < .001$) but not with the 10 percent condition (–0.28 point, $p = .33$). For anger, the effect also only holds for the 1 percent condition (–1.61 points, $p < .001$), while the 10 percent condition sees a slightly *higher* average level of anger than the tree episodic frames combined (–0.62 point, $p < .05$).

In summary, we can note that the episodic information induces stronger emotional responses on the *compassion* dimension than does *any* form of statistical information. However, for other emotional dimensions episodic information only has a stronger impact than the 1 percent statistical information condition (i.e., describing a single persons erroneous treatment). Overall, we find some support for our hypothesis that episodic information gives a stronger emotional response than statistical information (hypothesis 3).

Study 3: Immediate Recall of Statistical and Episodic Performance Information

In study 3, we turn to the question of differences in the recall of episodic and statistical information (hypothesis 4). Here we rely on a recall task in which subjects are randomly assigned to either recall some statistical or episodic information about public services.

Experimental Design and Procedure

The experiment is a two (episodic/statistical) x three (positive/mixed/negative) between-subjects design ($n = 674$). The wording of the treatments in the six groups is outlined in table 3. An example of one of the screens, as it was shown to the subjects, is reported in appendix C. In the episodic conditions, subjects are probed with concepts of "stories" or "experiences" with public services that are either positive, negative, or both (mixed). In the three statistical conditions, subjects are asked the same but with regard to "numbers and statistics." The split on negative, positive, and mixed recall is made in order to make the findings robust to systematic differences in the valence of the episodic and thematic information that citizens recall.

In the experiment, subjects are then asked to report what immediately comes to their mind in a large text box. Subjects have to report *something* in order to proceed in the survey. Open-ended questions are infrequently applied in survey experiments because they can be time-consuming to code and categorize the responses. However, the questions are well suited for recording what is on the subjects' minds without making cues in the response scale that implicitly link to treatments embedded in the question (RePass 1971). An alternative way of testing recall would have been to assign either episodic or statistical information to citizens at one point in time and then later directly ask them to recall the randomly assigned information (Aarøe and Petersen 2013).

In the analysis, we rely on two hand-coded measures of how citizens recall and report the information: First, an independent coder coded a simple dummy variable indicating whether subjects provided a response to the posed question in the treatment conditions. The coder was blinded to the treatment conditions and was only provided the open-ended responses in random order during the coding process. We thereby make sure that the coding is independent of knowledge about which treatment condition each response belongs to. The coding was done according to the following criteria: all responses containing "don't know" replies or other indications of inability to recall any information are coded as nonresponses. An example of the latter would be if subjects wrote a nonsensical string of letters or words. The majority of nonresponses was due to the subject writing "can't recall anything" or simply "nothing." If subjects provided some text that responded to the task, no matter how short, it was classified as a recall. Across all conditions the average share of recalls was 53.6 percent. Second, the

Table 3 Experimental Conditions

Treatment Conditions (randomly assigned):					
Episodic			Statistical		
Positive ($n = 107$)	Mixed ($n = 107$)	Negative ($n = 111$)	Positive ($n = 116$)	Mixed ($n = 118$)	Negative ($n = 115$)
There are many positive stories and experiences about how good public services are in Denmark.	*There are many positive and negative stories and experiences about how good or bad public services are in Denmark.*	*There are many negative stories and experiences about how bad public services are in Denmark.*	*There are many positive numbers and statistics about how good public services are in Denmark.*	*There are many positive and negative numbers and statistics about how good or bad public services are in Denmark.*	*There are many negative numbers and statistics about how bad public services are in Denmark.*
Describe the story or experience that immediately comes to mind here in the box below:			*Describe the number or statistic that immediately comes to mind here in the box below:*		

independent coder also provided an assessment of the "concreteness" and level of detail in each response that was coded. Concreteness is coded on an 11-point scale ranging from "not at all concrete" (0) to "very concrete" (10). Moreover, the coder was instructed to provide higher scores for responses that contain more details. The average concreteness for actual recalls across all conditions is 4.7 (SD = 2.1, $n = 361$). We apply the concreteness measure as a subjective indicator of how elaborate the recalled information is.

As a robustness check, we also rely on two indirect measures of the extent and nuance to the response: First, the number of words that each subject applies in their responses was counted. Across all conditions, respondents who actually recall information use on average 17.3 words and a median of 10 words. Word count is used as an alternative measure of how elaborate the responses are. This measure is validated by the fact that it is positively and significantly correlated with the subjectively coded concreteness score (adjusted $R^2 = 0.21$, $p < .001$). Second, we obtain the response time in seconds for each subject. Longer response times have been used as an indicator of more thorough processing (Petersen et al. 2013; Rubinstein 2013). The trimmed mean response time for valid responses is about 98 seconds.[7] Importantly, the measures capturing concreteness, number of words, and response time should be viewed as complementary and all serving the same purpose of testing whether episodic recall is more elaborate than recall of statistical information. If all measures show the same variation across the episodic recall conditions relative to the statistical recall conditions, this lends more support to hypothesis 4.

Table 4 Main Results

	Episodic	Statistical	*Difference*
Percent recalling information	63.2% ($n = 325$)	31.0% ($n = 349$)	32.1%*** [24.7%–39.6%]
Mean concreteness of response (0–10)	5.3 ($n = 228$)	3.8 ($n = 133$)	1.5*** [1.1–1.9]
Mean number of words in response	21.2 ($n = 228$)	10.8 ($n = 133$)	10.4*** [6.0–14.9]
Trimmed mean response latency (sec.)	112.0 ($n = 211$)	77.4 ($n = 116$)	34.6** [12.6–56.7]

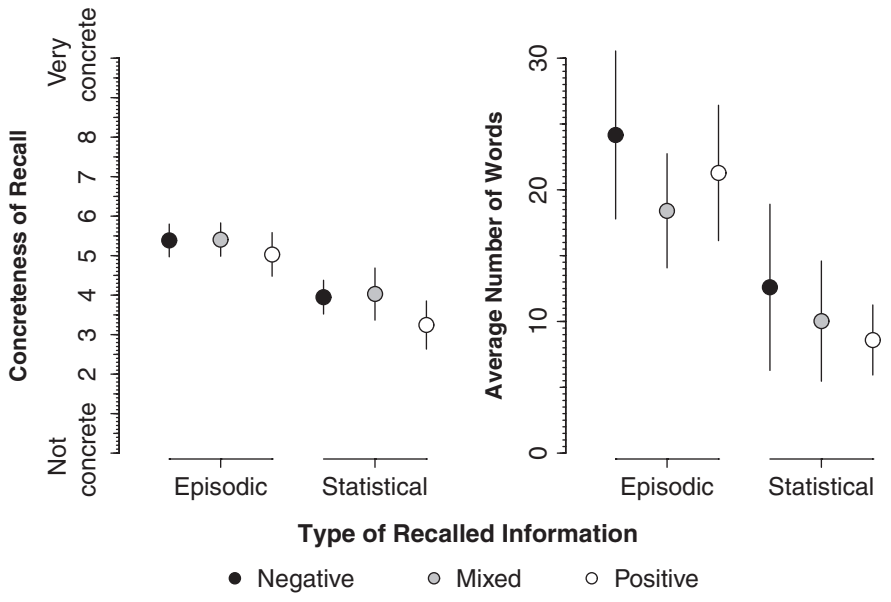
Note: Numbers in brackets are 95 percent confidence intervals.
*$p < .05$; **$p < .01$; ***$p < .001$.

Empirical Results

Results across the treatment groups are reported in table 4 and further outlined in figure 4. First, we look at the percent of respondents providing an answer to the inquiry in the treatment text. The results show that around twice as many citizens are able to recall information when asked about stories or experiences rather than when asked about statistics and numbers. The difference in reporting is highly significant and confirms the expectation that episodic performance information appears more likely to be recalled.[8]

Next, we examine how elaborate the responses are among those citizens who actually provide a response. The first indicator is the concreteness coding of the response. Our results show that episodic responses have more elaborate details than the statistical ones. On average, the concreteness of the episodic recall is 1.5 points higher, which is a substantial and significant difference on an 11-point scale. The average concreteness of responses across all six groups is plotted in figure 4 (left panel). Here, we can see that the difference in concreteness between episodic and statistical recall is very constant across negative, mixed, and positive responses.

The average number of words across all six groups is plotted in right panel in figure 4. In the episodic treatments, citizens use around



Notes: Left panel: Dots represent mean "concreteness" for responses in each condition. Right panel: Dots represent average number of words for responses in each condition. Both panels: Lines constitute 95 percent confidence intervals. Both panels only rely on content analysis of responses coded as an actual recall of information (***n*** = 361 in each panel).

Figure 4 Characteristics of the Recalled Information from Open-Ended Responses

20 words in their immediate recall of recent stories or experiences with public services. This is significantly more and about twice as many than citizens use when recalling statistics or numbers about public services. Again, we can see that the main difference in word usage between episodic and statistical recall is stable across negative, positive, and mixed categories. Finally, as a second objective indicator of thoughtful and elaborate responses, we can compare response latencies across groups for citizens who provided responses. These numbers confirm the impression from the concreteness coding and the word count: citizens asked to engage in episodic recall of public service performance take around 112 seconds (cf. table 4). This is much longer than those asked to retrieve statistics or numbers about public services who take around 77 seconds. In summary, we find a very consistent pattern where episodic information is more likely to be recalled, and the recall is more detailed and elaborate (hypothesis 4).

Discussion

Results from a survey experiment are limited in terms of their external validity across contexts and methods of delivering the treatment. However, it is not evident that the differences found in this study would be washed out in a real-world setting. In fact, in an information-rich environment, we would expect individuals to draw even more on vividness cues when sorting through hard and soft performance data. That being said, the experimental conditions applied in this study would be easy to use in a conceptual replication within an actual political-administrative setting. For instance, with field experiments, one could test how the very same treatment texts affect citizens when delivered as randomly assigned government reports with different degrees of episodic and statistical information. Another natural extension and test of external validity would be to validate other modes of communicating vivid performance information. To illustrate, both episodic and statistical information could be manipulated in terms of vividness by presenting data visualizations or photographs.

The results bring forth two important implications for the field of study centered on the effects of performance information: First, the results question the overall potency of statistical information to affect citizens' evaluations and recall of performance information relative to more idiosyncratic and personalized pieces of information. Numbers can no doubt be powerful and affect citizens' evaluations of public services. Still, the results confront practitioners with the daunting task of thinking carefully about how to engage citizens with the available performance data. As we have already noted, public administration scholars acknowledged the challenge of balancing "human interest" with "dull statistics" early on (Ridley 1937; Roher 1941; Upton 1915). For researchers, the results point to the importance of broadening the research program on how citizens draw inferences from performance information about public services. Here, the results tie into recent studies that stress how motivated reasoning and prior beliefs color citizens' interpretations of performance data (Bækgaard and Serritzlew 2016; Marvel 2016). With our findings on episodic information and the general importance of informational vividness, we have yet another obstacle for hard performance data to set its mark on the public at large. A natural extension of the findings presented here would be to focus on how individual differences affect the ways in which citizens draw on episodic and statistical information. For instance, numerical ability (numeracy) has been identified as an important moderator for how statistical information feeds into citizens' decisions (Peters et al. 2006). Furthermore, the results might add to our understanding of why research on citizen satisfaction often finds that citizens are less satisfied with performance at an abstract level than with specific services. One possibility is that citizens' reluctance to draw on statistics implies that general service impressions will be affected by unconscious biases against public services (Marvel 2016). Also, it helps explain why measures of citizen satisfaction often are weakly correlated with "hard," objective measures of performance (Stipak 1979): if citizens mostly rely on experiences and stories, it is more likely that their satisfaction with a service will diverge from the general pattern captured by a statistical measure of performance.

Another important implication of the findings is to understand their reach beyond the citizen level. For example, are employees or managers affected differently by episodic and statistical information than citizens are? Moynihan points to the potential strength of human-interest data at the managerial level: "Managers prefer to spend their time interacting with people and collecting oral data, not contemplating quantitative data" (2008, 167). Another recent observational study also points to the fact that mangers often draw on nonroutine episodic information (Kroll 2013). The obvious next step is to make a direct experimental comparison to how episodic and statistical information feeds into the decisions made by managers, employees, and organizations at large (Olsen 2015b).

Conclusion

For better or worse, most scholars agree that numerical performance information has a large impact on citizens' perceptions of public services. The findings of this article challenge this belief with evidence from a large-scale representative survey experiment. When asked to evaluate a public organization, citizens overwhelmingly choose statistical information over episodic information. Citizens' stated preferences are geared toward numerical performance information. Nonetheless, the subsequent findings highlight that the information that citizens *select* for performance evaluation is very different from the information that actually *affects* their evaluation and *recall* of public sector performance. In terms of impact on citizens' evaluations, we find that episodic information has a stronger effect when compared with statistical information that also highlights poor performance experienced by a single individual. Only statistical information that indicates severe negative performance can elicit a stronger negative response than a description of a single bad experience. These results echo findings in psychology on the importance of informational vividness and extend into the realm of performance data. The results also highlight how statistical information fails to induce an emotional response to the same degree that episodic information does. This outcome is particularly true for the emotional response of compassion. Finally,

In terms of impact on citizens' evaluations, we find that episodic information has a stronger effect when compared with statistical information that also highlights poor performance experienced by a single individual.

when asked to recall either individually based stories or statistical facts about the public sector, citizens find it much easier to recall the former than the latter. The difference is substantive and spans many dimensions. Citizens are much more able to describe stories and experiences with greater concreteness than statistics about public services.

As already noted, the obvious follow-up question is the scope of these findings beyond the citizen level. Managers and organizations often signal a great interest in hard performance data and spend many resources collecting quantitative measures. But perhaps managers are no different from the public at large: they state a preference for hard numerical data, while in reality, they allow their judgments, emotions, and memory to be swayed by the episodic information of everyday life.

Acknowledgments

The author thanks Carolyn J. Heinrich, Don Moynihan, Gregg Van Ryzin, Martin Vinæs Larsen, Nicolai Petrovsky, and Oliver James for their very helpful comments and suggestions for earlier drafts of the manuscript. The author also thanks three anonymous reviewers and the editorial team for extremely helpful comments. An earlier version was presented at the Public Management Research Conference in Minneapolis, Minnesota, June 11–13, 2015. The author would also like to thank participants at the panel for their helpful comments and suggestions.

Funding

The author is grateful for financial support from Grant no. 0602-02577B from the Danish Council for Independent Research and from a grant from the Master of Public Governance program at the University of Copenhagen under the project title "The Psychology of the Public Sector."

Notes

1. The use of numbers to impress and signal competence has also been recognized as a strategy of politicians in budgetary negotiations (Wildavsky 1964, 95).
2. Stone (2012) applies a similar dichotomy by contrasting "symbols" and "numbers." Here, symbols refer to stories and narratives.
3. A similar argument is stressed in the availability bias, which states that individuals draw more heavily on exemplars that easily come to mind (Taylor and Thomson 1982).
4. In a review, Taylor and Thomson summarize this fundamental difference as follows: "Case histories are assumed to be more vivid than their dry statistical counterparts (base-rate information) by virtue of the fact that they contain more concrete information and more imaginable material than does base-rate information. They may also arouse more of an emotional response in the perceiver and, for all three of the above reasons, may have a stronger representation in memory" (1982, 162).
5. Participating subjects received points, which can be used for purchases and lotteries conducted by YouGov.
6. A total of 3,502 subjects were invited to participate in the survey, yielding a response rate of 28.9 percent fully completed surveys.
7. We rely on trimmed means where the top and bottom 5 percent are removed. It is standard practice to rely on trimmed means, medians, or some other transformation for the response time in order to account for some subjects spending very long amounts of time (Rubinstein 2013). This could happen if respondents were to pause in the middle of the survey.
8. An example of a typical recalled *story* (translated from Danish) is as follows: "A demented female wheelchair user from Svendborg who could not get any pension, when the municipality did not think she was ill enough to get any pension. Now she is dead." An example of a typical recalled *statistic* would be "treatment time in health care."

References

Aarøe, Lene. 2011. Investigating Frame Strength: The Case of Episodic and Thematic Frames. *Political Communication* 28(2): 207–26.

Aarøe, Lene, and Michael Bang Petersen. 2013. Episodic Bias and the Social Transmission of Political Communication. Paper presented at the Annual Meeting of the Danish Political Science Association. Demark: Vejle.

Bækgaard, Martin. 2015. Performance Information and Citizen Service Attitudes: Do Cost Information and Service Use Affect the Relationship? *International Public Management Journal* 18(2): 228–45.

Bækgaard, Martin, and Søren Serritzlew. 2016. Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension. *Public Administration Review* 76(1): 73–82.

Bolman, Lee G., and Terrence E. Deal. 1991. *Reframing Organizations: Artistry, Choice, and Leadership*. San Francisco: Jossey-Bass.

Charbonneau, Étienne, and Gregg G. Van Ryzin. 2015. Benchmarks and Citizen Judgments of Local Government Performance: Findings from a Survey Experiment. *Public Management Review* 17(2): 288–304.

Daschmann, Gregor. 2000. Vox Pop and Polls: The Impact of Poll Results and Voter Statements in the Media on the Perception of a Climate of Opinion. *International Journal of Public Opinion Research* 12(2): 160–81.

Davies, Ian C. 1999. Evaluation and Performance Management in Government. *Evaluation* 5(2): 150–59.

Espeland, Wendy N., and Michael Sauder. 2007. Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology* 133(1): 1–40.

Gross, Kimberly. 2008. Framing Persuasive Appeals: Episodic and Thematic Framing, Emotional Response, and Policy Opinion. *Political Psychology* 29(2): 169–92.

Grosso, Ashley L., and Gregg G. Van Ryzin. 2011. How Citizens View Government Performance Reporting: Results of a National Survey. *Public Performance and Management Review* 35(2): 235–50.

Hansen, Kasper M., Asmus Leth Olsen, and Mickael Bech. 2015. Cross-National Yardstick Comparisons: A Choice Experiment on a Forgotten Voter Heuristic. *Political Behavior* 37(4): 767–89.

Herbst, Susan. 1993. *Numbered Voices: How Opinion Polling Has Shaped American Politics*. Chicago: University of Chicago Press.

Herr, Paul M., Frank R. Kardes, and John Kim. 1991. Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research* 17(4): 454–62.

Holzer, Marc, and Kaifeng Yang. 2004. Performance Measurement and Improvement: An Assessment of the State of the Art. *International Review of Administrative Sciences* 70(1): 15–31.

Hood, Christopher. 2007. Public Service Management by Numbers: Why Does It Vary? Where Has It Come From? What Are the Gaps and the Puzzles? *Public Money and Management* 27(2): 95–102.

Iyengar, Shanto. 1990. Framing Responsibility for Political Issues: The Case of Poverty. *Political Behavior* 12(1): 19–40.

Jackson, Peter M. 2011. Governance by Numbers: What Have We Learned over the Past 30 Years? *Public Money and Management* 31(1): 13–26.

James, Oliver. 2011. Performance Measures and Democracy: Information Effects on Citizens in Field and Laboratory Experiments. *Journal of Public Administration Research and Theory* 21(3): 399–418.

James, Oliver, and Alice Moseley. 2014. Does Performance Information about Public Services Affect Citizens' Perceptions, Satisfaction, and Voice Behaviour?

Field Experiments with Absolute and Relative Performance Information. *Public Administration* 92(2): 493–511.

James, Oliver, and Gregg G. Van Ryzin. 2015. Incredibly Good Performance: An Experimental Study of Source and Level Effects on the Credibility of Government. *American Review of Public Administration.* Published electronically on April 9. doi:10.1177/0275074015580390.

Jenni, Karen E., and George Loewenstein. 1997. Explaining the "Identifiable Victim Effect." *Journal of Risk and Uncertainty* 14(3): 235–57.

Kettl, Donald F. 2016. Making Data Speak: Lessons for Using Numbers for Solving Public Policy Puzzles. *Governance.* Published electronically on May 20. doi:10.1111/gove.12211.

Kogut, Tehila, and Ilana Ritov. 2005. The "Identified Victim" Effect: An Identified Group, or Just a Single Individual? *Journal of Behavioral Decision Making* 18(3): 157–67.

Kroll, Alexander. 2013. The Other Type of Performance Information: Non-Routine Feedback, Its Relevance and Use. *Public Administration Review* 73(2): 265–76.

Lee, Mordecai. 2006. Empirical Experiments in Public Reporting: Reconstructing the Results of Survey Research, 1941–1942. *Public Administration Review* 66(2): 252–62.

March, James G., and Herbert A. Simon. 1958. *Organizations*. New York: Wiley.

Marvel, John D. 2016. Unconscious Bias in Citizens' Evaluations of Public Sector Performance. *Journal of Public Administration Research and Theory* 26(1): 143–58.

Mintzberg, Henry. 1971. Managerial Work: Analysis from Observation. *Management Science* 18(2): 97–110.

_____. 1973. *The Nature of Managerial Work*. New York: Harper & Row.

Moynihan, Donald P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.

Nisbett, Richard E., and Ross Lee. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice Hall.

Olsen, Asmus Leth. 2013. Leftmost-Digit-Bias in an Enumerated Public Sector? An Experiment on Citizens Judgment of Performance Information. *Judgment and Decision Making* 8(3): 365–71.

_____. 2015a. Citizen (Dis)satisfaction: An Equivalence Framing Study. *Public Administration Review* 75(3): 469–78.

_____. 2015b. The Numerical Psychology of Performance Information—Implications for Citizens, Managers, and Policy Makers. *Public Performance and Management Review* 39(1): 100–115.

_____. Forthcoming. Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations. *Journal of Public Administration Research and Theory*.

Osborne, David, and Ted Gaebler. 1992. *Reinventing Government: How the Entrepreneurial Spirit Is Transforming Government*. Reading, MA: Addison-Wesley.

Perrin, Burt. 1998. Effective Use and Misuse of Performance Measurement. *American Journal of Evaluation* 19(3): 367–79.

Peters, Ellen, Daniel Västfjäll, Slovic Paul, C.K. Mertz, Ketti Mazzocco, and Stephan Dickert. 2006. Numeracy and Decision Making. *Psychological Science* 17(5): 407–13.

Petersen, Michael Bang, Martin Skov, Søren Serritzlew, and Thomas Ramsøy. 2013. Motivated Reasoning and Political Parties: Evidence for Increased Processing in the Face of Party Cues. *Political Behavior* 35(4): 831–54.

Pettus, Clinton, and Edward Diener. 1977. Factors Affecting the Effectiveness of Abstract Versus Concrete Information. *Journal of Social Psychology* 103(2): 233–42.

Radin, Beryl. 2006. *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington, DC: Georgetown University Press.

RePass, David E. 1971. Issue Salience and Party Choice. *American Political Science Review* 65(2): 389–400.

Ridley, Clarence E. 1937. Municipal Reporting Taken Seriously. *Public Opinion Quarterly* 1(1): 112–16.

Roher, Miriam. 1941. Education of a Citizen. *National Municipal Review* 30(4): 192–98.

Rubinstein, Ariel. 2013. Response Time and Decision Making: An Experimental Study. *Judgment and Decision Making* 8(5): 540–51.

Slattery, Karen L., and Ernest A. Hakanen. 1994. Sensationalism versus Public Affairs Content of Local TV News: Pennsylvania Revisited. *Journal of Broadcasting and Electronic Media* 38(2): 205–16.

Slovic, Paul. 2007. If I Look at the Mass I Will Never Act: Psychic Numbing and Genocide. *Judgment and Decision Making* 2(2): 79–95.

Smith, Peter. 1995. On the Unintended Consequences of Publishing Performance Data in the Public Sector. *International Journal of Public Administration* 18(2): 277–310.

Stipak, Brian. 1979. Citizen Satisfaction with Urban Services: Potential Misuse as a Performance Indicator. *Public Administration Review* 39(1): 46–52.

Stone, Deborah A. 2012. *Policy Paradox: The Art of Political Decision Making*, Rev. ed. New York: W. W. Norton.

Taylor, Shelley E., and Suzanne C. Thompson. 1982. Stalking the Elusive Vividness Effect. *Psychological Review* 89(2): 155–81.

Tulving, Endel. 2002. Episodic Memory: From Mind to Brain. *Annual Review of Psychology* 53: 1–25.

Upson, Lent D. 1915. The Value of Municipal Exhibits. *National Municipal Review* 4(1): 65–69.

Van Dooren, Wouter, Geert Bouckaert, and John Halligan. 2010. *Performance Management in the Public Sector*. New York: Routledge.

Wildavsky, Aaron B. 1964. *Politics of the Budgetary Process*. Boston: Little, Brown.

Zhang, Y. Charles, and Norbert Schwarz. 2013. The Power of Precise Numbers: A Conversational Logic Analysis. *Journal of Experimental Social Psychology* 49(5): 944–46.

Zillmann, Dolf. 2006. Exemplification Effects in the Promotion of Safety and Health. *Journal of Communication* 56(1): 221–37.

Appendix A. Study 1: Example of a Treatment Condition as Viewed By the Subjects in the Survey

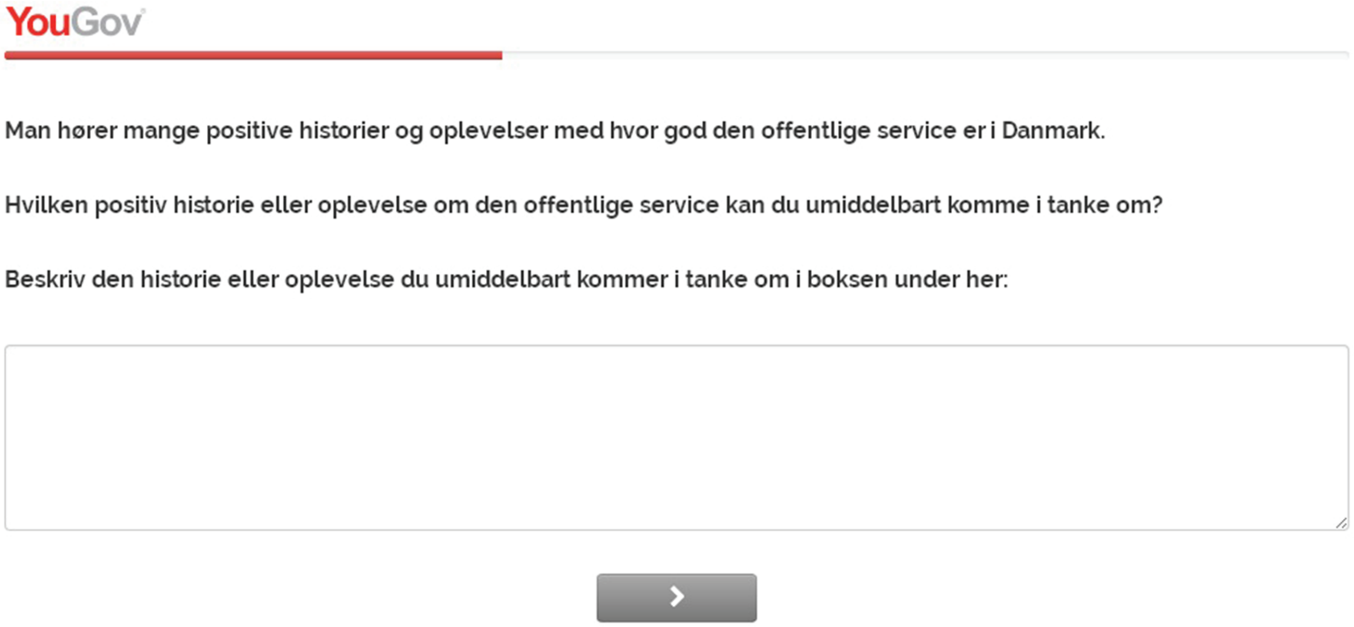
Note: Statistics versus episodic case of "Birgitte." "*Imagine that you have to evaluate how good or bad a hospital is. You can choose one of the two following pieces of information. What will you choose? ( ) Results from a satisfaction report with 100 former patients. ( ) The experience that 51-year-old Birgitte had as a patient.*"

Appendix B. Study 2: Example of a Treatment Condition as Viewed By the Subjects in the Survey

Note: Episodic case "Erik." "*Imagine that you know the following about a hospital: 31-year-old Erik was admitted to the hospital last year in order to have a simple surgery done. Following the surgery, Erik was in great pain due to erroneous treatment. How good do you think the hospital is?*"

Note: Emotional response. "*To what extent does this information make you feel: anger, compassion, disgust.*"

Appendix C. Study 3: Example of a Treatment Condition as Viewed by the Subjects in the Survey



Note: Episodic recall—positive. *"There are many positive stories and experiences with how good public services are in Denmark. Describe the story or experience that immediately comes to mind in the box below here:"*